Using Reality Mining to Improve Public Health and Medicine

Alex (Sandy) Pentland, David Lazer, Devon Brewer, Tracy Heibeck
pentland@mit.edu, david_lazer@harvard.edu, dbrewer@interscientific.net, theibeck@hotmail.com

***Abstract*—Digital networks not only connect us, they sense where we go and with whom we communicate. Statistical analysis of these digital records, a process called "reality mining," allows us to create a surprisingly comprehensive picture of our individual and collective lives. Computational models based on such data could transform many areas of human life, including improvements in public health and medicine. Current legal statutes are lagging far behind our data collection capabilities, making it crucial to begin discussing how this technology will and should be used.**

Keywords-reality mining; public health; social networks; social behavior pattern

I. Introduction

We live our lives in digital networks. We wake up in the morning, check our e-mail, make a quick phone call, commute to work, buy lunch. Many of these transactions leave digital breadcrumbs – tiny records of our daily experiences. Reality mining, which pulls together these crumbs using statistical analysis and machine learning methods, offers an increasingly comprehensive picture of our lives, both individually and collectively. Due to its potential to transform our understanding of ourselves, our organizations, and our society in a fashion that was barely conceivable just a few years ago, reality mining was recently identified by *Technology Review* as one of "10 emerging technologies that could change the world" [1].

Many everyday devices provide the raw database upon which reality mining builds; sensors in mobile phones, cars, security cameras, RFID ('smart card') readers, and others, all allow for the measurement of human physical and social activity. Currently, the single most important source of reality mining data is the ubiquitous mobile phone. Every time a person uses a mobile phone, a few bits of information are left behind. The phone pings the nearest mobile-phone towers, revealing its location. Accelerometers already found in some phones can record patterns of physical activity, and the phone's signal processing hardware can analyze its user's speaking patterns. With the aid of data-mining algorithms, these data could shed light on individual patterns of behavior and even on the well-being of communities, creating new ways to improve health and medicine.

Within the next few years reality mining will become more common, thanks in part to the proliferation and increasing sophistication of mobile phones. Many handheld devices now have the processing power of low-end desktop computers, and they can also collect more varied data, due to components such as GPS chips that track location. The Chief Technology Officer of EMC, a large digital storage company, estimates that this sort of personal sensor data will balloon from 10% of all stored information to 90% within the next decade.

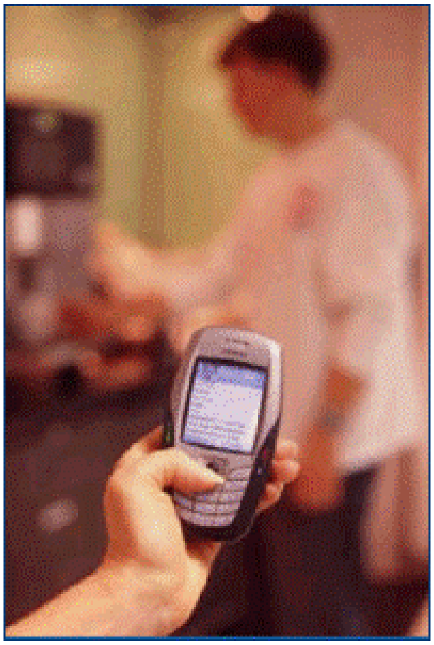
While the promise of reality mining is great, the idea of collecting so much personal information naturally raises many questions about privacy. It is crucial that behavior-logging technology not be forced on anyone. Legal statutes are lagging woefully behind our rapidly expanding data collection capabilities, however, making it crucial to begin discussing how this technology will and should be used.

II. Mapping Social Networks

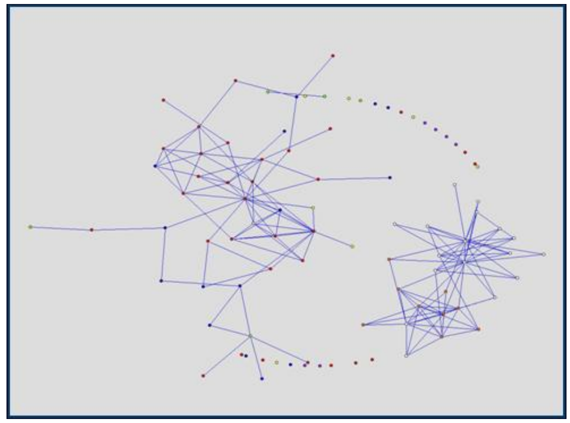
One of the most important applications of reality mining may be the automatic mapping of social networks [2]. In Figure 1(a), you see a smart phone that is programmed to sense and report continuously on its user's location, who else is nearby, the user's call and SMS patterns, and (with phones that have accelerometers) how the user is moving. One hundred of these phones were deployed to students at MIT during the 2004-2005 academic year. Figure 1(b) shows the patterns of proximity among the participants during one day; even casual examination shows that the students were part of two separate groups: the Sloan School and the Media Lab.

Careful analysis of these data shows different patterns of behavior depending upon the social relationship between people. Figure 1(c) shows the pattern of proximity during one week, and it can be seen that self-reported reciprocal friends (both persons report the other as a friend), non-reciprocal friends (only one of a pair reports the other as a friend), and reciprocal non-friends (neither of a pair reports the other as a friend) exhibit very different patterns [3]. By using more sophisticated statistical analysis, we can map each participant's social network of friends and co-workers with an average accuracy of 96% [4].

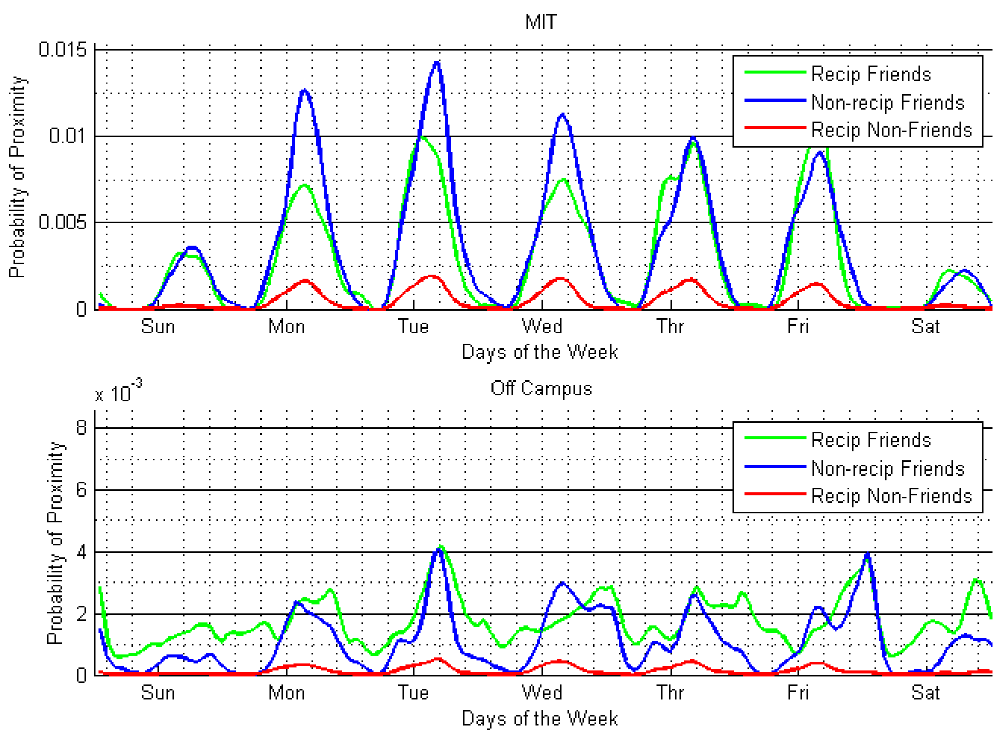
This paper is based on a February 2009 whitepaper commissioned by the Robert Wood Johnson Foundation.



A



B



C

Figure 1: *Mapping social networks from mobile phone location/proximity data. 1(a) shows a 'smart phone' programmed to sense other people using Bluetooth, 1(b) shows the pattern of proximity between people during one day, and 1(c) shows that different social relationships are associated with different patterns of proximity.*

Figure 2: *Analysis of travel patterns allows discovery of largely independent subpopulations within a city. Movement patterns (a), measured from GPS mobile phones, allow (b) segmentation of the population into subpopulations with differing behavior patterns, and measurement of the 'mixing' between those groups [7].*

Reality mining's capability for automatic social network mapping is now being used in a variety of research applications. As an example, a current research project underway at MIT is aimed at understanding health-related behaviors and infectious disease propagation. At this time, we have above 80% participation of students in a MIT dormitory that includes freshmen and upperclassmen, and are beginning to compare the behavior and health changes that freshmen normally experience with the changes in their various social networks. This experiment should help to disentangle causal pathways about how social networks influence obesity and other health-related behaviors, as well as provide unprecedented detail for modeling the spread of infectious disease.

III. BEYOND DEMOGRAPHICS TO SOCIAL BEHAVIOR PATTERNS

Most government health services rely on demographic data to guide service delivery. Demographic characteristics, however, are a relatively poor predictor of individual behavior, and it is behavior – not wealth, age, or place of residence – that is the major determinant of many health outcomes. Reality mining provides a way to characterize behavior, and thus provides a classification framework that is more directly relevant to health outcomes [5].

The pattern of movement between the places a person lives, eats, works, and hangs out are known as a *social behavior pattern.* Reality mining research has shown that most people have only a small repertoire of these social behavior patterns, and that this small set of social behavior patterns accounts for the vast majority of an individual's activity [6].

The fact that all mobile phones constantly measure their position (either through GPS or by finding the nearest cell tower) means that we can use reality mining of mobile phone location data to directly characterize an individual's set of social behavior patterns. We can also cluster together people with similar social behavior patterns in order to discover the independent subgroups within a population.

Figure 2(a) shows movement patterns with popular 'hang outs' color coded by the different subpopulations that populate these destinations, where the subpopulations are defined by both their demographics and, more importantly, by their *behaviors.* Figure 2(b) shows that the mixing between these different behavior subpopulations is surprisingly small.

Understanding the social behavior patterns of different subpopulations and the mixing between them is critical to the delivery of public health services, because different subpopulations have different risk profiles and different attitudes about health-related choices. The use of reality mining to discover these social behavior patterns can potentially provide great improvements in health education efforts and behavioral interventions.

IV. THE FUTURE POTENTIAL OF REALITY MINING: IMPROVING PUBLIC HEALTH

In the previous sections, we discussed how reality mining has the potential to map social networks automatically and to discover subpopulations with different social behavior patterns. By leveraging this new understanding of social networks, we may achieve significant advances in terms of behavioral change and health.

For example, research suggests that some chronic health-related conditions/behaviors are "contagious," in the sense that individual-level outcomes are linked to other individuals with whom one shares social connections. Both smoking behavior [8] and obesity [9] seem to spread within social networks. Smoking and obesity likely serve as good models for other health related behaviors, such as diet, exercise, general hygiene, and so on.

These findings, however, beg for an examination of the causal mechanism – an essential step if interventions are to be designed to improve public health. For example, is the diffusion of these behaviors and conditions driven by the emergence of norms within the network – e.g., smoking is cool; one should exercise frequently, etc.? Alternatively, is the diffusion driven directly by the social component of the relevant behaviors – e.g., smoking, eating, or exercising with one's friends? Or might the apparent spread of these behaviors reflect individuals seeking out others with similar inclinations? The type of data needed to understand the causal mechanism is exactly the fine granularity data that reality mining can provide.

Another potentially important application of reality mining is the tracking of infectious diseases. As the world becomes increasingly interconnected through the movement of people and goods, the potential for global pandemics of infectious disease rises as well. In recent years, outbreaks of SARS and other serious infectious diseases in widely separated but socially linked communities highlight the need for fundamental research on disease transmission and effective prevention and control strategies.

With GPS and related technologies, it is increasingly easy to track the movements of people [10, 2]. Logs of location tracking data from cell phones could prove invaluable to public health officials when investigating cases of serious infectious disease (e.g., tuberculosis, SARS, measles, Legionnaires' disease, etc.) to help identify the source of infections and prevent further transmission. People often forget all the locations they have visited, even for recent periods, and similarly might not know many of the people to whom they were exposed or might have exposed themselves. All of this underlines the potential value of systematically analyzing such records for disease control.

Reality mining, although still in its infancy, is poised to quickly become more common, due in large part to the rapid proliferation and increasing sophistication of mobile phones. Many mobile phones and other technologies already collect a great deal of information about their users – data such as physical activity and conversational cadences – and this will only increase. In the near future it may be common for smart phones to continuously monitor a person's motor activity, social interactions, sleep patterns, and other health indicators. Computational models based on such data could dramatically transform many areas of human life, including the possibility of significant improvements in public health and medicine. While such data pose a potential threat to individual privacy, they also offer great potential value both to individuals and communities. Current legal statutes are lagging far behind our data collection capabilities, making it particularly important to begin discussing how this technology will and should be used.

REFERENCES

[1] K. Greene, "Emerging technologies 2008: Reality mining," Technology Reivew, vol. 111(2), pp. 54-56, April 2008.

[2] N. Eagle and A. Pentland, "Reality mining: Sensing complex social systems," Personal and Ubiquitous Computing, vol. 10(4), pp. 255-268, 2006.

[3] N. Eagle, D. Lazer, and A. Pentland, "Inferring friendship from proximity," to appear PNAS, Sept. 2009.

[4] W. Dong and A. Pentland, "Modeling influence between experts," Lecture Notes on AI: Special Volume on Human Computing, vol. 4451, pp. 170-189, 2007.

[5] A. Pentland and T. Heibeck, Honest Signals: How They Shape Your World. Cambridge, MA: MIT Press, 2008.

[6] A. Pentland, "Automatic mapping and modeling of human networks," Physica A: Statistical Mechanics and its Applications, vol. 378(1), pp. 59-67, 2007.

[7] Sense Networks, http://www.sensenetworks.com, 2008.

[8] N. Christakis and J. Fowler, "The collective dynamics of smoking in a large social network," New England Journal of Medicine, vol. 358, pp. 2249-2258, 2008.

[9] N. Christakis and J. Fowler, "The spread of obesity in a large social network over 32 years," New England Journal of Medicine, vol. 357, pp. 370-379, 2007.

[10] M. Gonzalez, C. Hidalgo, and A.L. Barabasi, "Understanding individual human mobility patterns," Nature, vol. 453, pp. 779-782, June 2008.